

An Overview of Systematic Reviews on Prognostic Factors in Neck Pain: Results from the International Collaboration on Neck Pain (ICON) Project

David M. Walton^{*1}, Linda J. Carroll², Helge Kasch³, Michele Sterling⁴, Arianne P. Verhagen⁵, Joy C. MacDermid⁶, Anita Gross⁶, P. Lina Santaguida⁶, Lisa Carlesso⁷ and ICON⁸

¹Faculty of Health Sciences, The University of Western Ontario, London Ontario, Canada

²Department of Public Health Sciences and Alberta Centre for Injury Control Research, School of Public Health, University of Alberta, Edmonton Alberta, Canada

³The Danish Pain Research Center, Department of Neurology, Aarhus University Hospital, Denmark

⁴Centre for National Research on Disability and Rehabilitation Medicine (CONROD), The University of Queensland, Brisbane Australia

⁵Department of General Practice, Erasmus Medical Centre University, Rotterdam, The Netherlands

⁶School of Rehabilitation Science, McMaster University, Hamilton Ontario, Canada

⁷Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton Ontario, Canada

⁸International Collaboration on Neck (ICON)

Abstract: Given the challenges of chronic musculoskeletal pain and disability, establishing a clear prognosis in the acute stage has become increasingly recognized as a valuable approach to mitigate chronic problems. Neck pain represents a condition that is common, potentially disabling, and has a high rate of transition to chronic or persistent problems. As a field of research, prognosis in neck pain has stimulated several empirical primary research papers, and a number of systematic reviews. As part of the International Consensus on Neck (ICON) project, we sought to establish the general state of knowledge in the area through a structured, systematic review of systematic reviews (overview).

An exhaustive search strategy was created and employed to identify the 13 systematic reviews (SRs) that served as the primary data sources for this overview. A decision algorithm for data synthesis, which incorporated currency of the SR, risk of bias assessment of the SRs using AMSTAR scoring and consistency of findings across SRs, determined the level of confidence in the risk profile of 133 different variables. The results provide high confidence that baseline neck pain intensity and baseline disability have a strong association with outcome, while angular deformities of the neck and parameters of the initiating trauma have no effect on outcome. A vast number of predictors provide low or very low confidence or inconclusive results, suggesting there is still much work to be done in this field. Despite the presence of multiple SR and this overview, there is insufficient evidence to make firm conclusions on many potential prognostic variables. This study demonstrates the challenges in conducting overviews on prognosis where clear synthesis criteria and a lack of specifics of primary data in SR are barriers.

Keywords: Overview, prognosis, whiplash, non-specific neck pain, work-related neck pain.

INTRODUCTION

Neck pain is one of the most common musculoskeletal disorders in the general population, with a 1-year point prevalence of approximately one-third of adults [1]. The severity of pain can range from minor to severely debilitating [2]. Effective management of neck pain requires knowledge of the best evidence for each of assessment, prognosis, intervention and outcome measurement. While a number of systematic reviews (SRs) have been published in each of these 4 domains, it is not uncommon that SRs reach

different conclusions when compared against one another. Differences in search strategy, range of publication date, quality scoring, data extracted and synthesis technique may explain the disparate conclusions. As part of a larger initiative to establish clear, actionable messages for all elements of neck pain management, the International Collaboration on Neck (ICON) group has performed a systematic review and synthesis of these SRs in an effort to identify consistent messages across diverse research groups. This manuscript will describe the findings from the prognosis arm of this initiative.

Prognosis is an important component of clinical decision making for any condition. Those with a positive prognosis may rarely require intervention beyond standard advice and education. Those with a poor prognosis should arguably be

*Address correspondence to this author at the Western University Canada, Rm. EC1443, School of Physical Therapy, 1201 Western Rd., London, ON N6H 1A1, Canada; Tel: 519-661-2111, Ext. 80145; Fax: 519-661-3866; E-mail: dwalton5@uwo.ca

considered for more in depth evaluation and targeted intervention in the early stages of the condition to prevent a transition to chronicity. Prognosis in neck pain is especially important; it has recently been estimated that approximately one half of all people with acute traumatic neck pain (eg. whiplash) will recover regardless of the intervention, while the other half will experience delayed recovery or chronic problems [3]. Without sound understanding of important prognostic factors, the decision of whether to initiate early targeted treatment or take a 'wait-and-see' approach will only be accurate 50% of the time. There have been a series of SRs and data syntheses published for both traumatic [3-9] and non-traumatic [10, 11] neck pain. The most consistent findings amongst these SRs are the poor prognostic value of high initial pain intensity or high aggregate scores on self-reported disability, which individually provide little guidance for intervention decisions.

An overview of systematic reviews (OvR) is a relatively new approach to synthesizing a large body of literature in an area. The approach requires similar search strategies and quality scoring as a systematic review of primary literature, but relies on the appraisal and data extraction of previous reviewers rather than going back to the primary sources. While this exposes the results of the OvR to potential bias, in the case where appraisal and extraction of the primary sources by previous reviewers was less than optimal, the inclusion and scoring of only peer-reviewed published SRs and a focus on temporal proximity and consistency provides acceptable confidence in the synthesis and results. The primary outcome in the prognostic SR was consistency of findings, with strength of the association between risk factor and outcome reported where available. Consistent findings, especially across recent high-quality SRs, provide confidence in the value of a risk factor. Pooled effect sizes are best left to targeted meta-analyses on the subject (e.g. [6, 9]).

The purpose of this overview of SRs was to identify consistent risk factors for delayed or non-recovery (i.e. chronic pain and/or disability) from neck pain, through a systematic process of searching, filtering, scoring and extracting results of published SRs of prognostic factors.

METHODS

Study Selection

The methods consistent across all 4 OvRs in the ICON project have been detailed in a previous paper in this series [Refer to Methods paper in this series]. Specific to the prognosis domain, a search strategy was constructed and applied to the following international databases searched from January 2000 to March 2012: MEDLINE, EMBASE, CINAHL, ILC, CENTRAL, and LILACS. Only peer-reviewed systematic SRs or meta-analyses were considered eligible for inclusion in the prognosis overview. Neither narrative, non-systematic reviews nor editorials/commentaries were eligible for this overview. Manuscripts were accepted if they were written in English, included only primary sources that focused on adults with neck pain of any cause, and evaluated prognostic factors for outcome of a current episode of neck pain. SRs identifying primary risk factors for the development of new onset neck pain in

otherwise healthy populations were not included in this prognosis overview. Potentially eligible SRs were screened by two independent screeners, first at the level of title and abstract, and then full text.

Quality Appraisal

Risk of bias appraisal and scoring was performed using the AMSTAR review methodologic checklist by two raters [12]. The AMSTAR provides a risk of bias assessment through 11 different domains. It has been shown to be adequately valid and reliable for use in assessing systematic reviews [13, 14]. Differences were settled through consensus. The determination of quality was made on a review-by-review basis, recognizing that some AMSTAR items are more relevant for determining the quality of a systematic review of prognostic factors than were others. Table 1 describes the AMSTAR items that were considered most relevant (and hence most highly weighted) for determining risk of bias. This individualized approach to establishing quality is consistent with current recommendations [15]. SRs were categorized as high (low risk of bias), medium (moderate risk of bias), or low (high risk of bias) quality based on this process.

Data Extraction

Results were extracted as described verbatim in each SR and compared by two independent reviewers. Prognostic factors were grouped by conceptual category into: Event-related (ie. parameters of the trauma), Psychological & Behavioural, Symptoms & Interference, Biological or Clinical Assessment, the Medicolegal context, Demographics, Other Social Factors, Pre-injury History, and Treatment-related. We retained the summarized level of evidence as described verbatim in each paper and entered that into a database. Some SRs reported summarized results as strong, moderate, limited or inconclusive evidence [5,7,8,11]. Other summary structures were similar, including 'consistent, inconsistent or inconclusive'[4] or 'consistent, balanced or limited' [16]. One meta-analysis was identified [6] that used a homogeneous subsample of the literature to present strength of the evidence based on both pooled effect size and fail-safe N. During the course of this overview, this meta-analysis was updated with a new literature search and the effects of 13 variables were adjusted based on new data. In the interest of being as current as possible, the revised effect sizes were included in this overview where available. Other SRs did not present summary levels of evidence, rather presenting the numbers of primary papers supporting or refuting each predictor [3, 9, 10, 17] supplemented in some cases by qualitative interpretation of the methodological rigour of each primary source [3, 9, 10, 17]. For these papers, an algorithm for determining level of evidence that would allow comparison with other SRs was created based on the consistency of findings. Strong evidence required at least 3 primary sources, with at least 2/3 finding similar results, moderate evidence required similar results in only 2 primary sources with no conflicting sources, limited evidence was present when only a single primary source was reviewed, and inconclusive evidence was present when less than 2/3 of the primary sources found similar results regardless of the absolute number.

Table 1. Included Systematic Reviews and Results for Each of the Relevant AMSTAR Indicators. Where an Item was Unclear in the Text, it was Marked as a 'no' in the AMSTAR Database

	Was there Duplicate Study Selection and Data Extraction?	Was a Comprehensive Literature Search Performed?	Was a List of Studies (Included and Excluded) Provided?	Where the Characteristics of the Included Studies Provided?	Was the Scientific Quality of the Included Studies Assessed and Documented?	Was the Scientific Quality of the Included Studies Used Appropriately in Formulating Conclusions?	Were the Methods Used to Combine the Findings of Studies Appropriate?	Was the Likelihood of Publication Bias Assessed?	Was the Conflict of Interest for All Included Studies Stated?	Was there Duplicate Study Selection and Data Extraction?	Was a Comprehensive Literature Search Performed?
Campbell 2011	N	Y	N	Y	Y	Y	n/a	N	N	N	Y
Carroll 2008a	Y	Y	N	Y	Y	Y	n/a	N	N	N	N
Carroll 2008b	Y	Y	N	Y	Y	Y	n/a	N	N	N	N
Carroll 2009	Y	Y	N	Y	Y	Y	n/a	N	N	N	N
Cassidy 2008	Y	N	N	N	Y	N	n/a	N	N	Y	N
Hush 2011	Y	Y	N	Y	Y	Y	Y	N	N	Y	Y
Kamper 2008	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y
McClune 2002	N	Y	N	N	N	N	n/a	N	N	N	Y
McLean 2007	Y	Y	N	Y	Y	Y	n/a	N	N	Y	Y
Scholten-Peeters 2003	Y	Y	N	Y	Y	Y	n/a	N	N	Y	Y
Walton 2009 ¹	N	Y	N	Y	Y	Y	Y	Y	N	N	Y
Williams 2007	Y	Y	N	Y	Y	Y	n/a	N	N	Y	Y
Williamson 2008	Y	Y	N	Y	Y	Y	n/a	N	N	Y	Y

¹Walton and colleagues updated their meta-analysis during the course of this overview but the update was pending publication. We have indicated the date of publication of their first meta-analysis, but have used results from the updated one where applicable. The scores on the AMSTAR tool are related to the original 2009 publication.

Analysis

In order to establish summarized findings, we considered both age and risk of bias (methodologic quality) of the SR. Recognizing the short shelf-life of SRs, Whitlock [15] suggests that greater weight should be given to more recent SRs, with older SRs providing supporting evidence only. Since effect size was rarely reported, the outcome of interest was limited to confidence in the existence and direction of an association between a predictor and a subsequent outcome (risk of poor outcome, no association with outcome, or inconclusive). In our case, confidence in the direction of each predictor was established through first evaluating the findings from the most recent SR(s) of at least medium quality. Where multiple SRs were published on the same topic within a relatively short time span, confidence in the conclusions regarding the direction and significance of effect for each predictor was an amalgam of 1) SR quality and 2) consistency in findings across different authorship groups. For example, during the years 2007-2009, 5 SRs on prognosis following whiplash were published [3, 6-9]. In light of the different methodologies for searching and

synthesizing results across the included SRs, our consistency approach can be considered analogous to triangulation for establishing trustworthy results in qualitative research [18].

Given the phrasing of each prognostic factor, in only one case was a factor described as protective (i.e. facilitate recovery): regular physical activity in the case of non-traumatic neck pain. The confidence in each association was categorized using an approach adapted from the GRADE working group [19]: High, moderate, low or very low confidence that the direction of association is robust to findings in future research. In an attempt to be conservative, high confidence was reserved for only those predictors for which consistent high-quality evidence was presented in each SR with at least 1 high quality SR and no conflicting SRs. Moderate confidence required consistent high-level findings from at least 1 recent medium-quality SR, with the majority of findings from other concurrent SRs (where applicable) in the same direction of effect. Low confidence was assigned to a predictor when summary findings were of low-moderate level from the majority of SRs with some conflicting results, or when only a single SR reported

significant but moderate findings for that predictor. Very low confidence was assigned when none of the above conditions were met. As a result of these algorithms, each predictor received both an estimate of its association with outcome (risk of poor outcome, no effect on outcome, inconclusive effect) and a level of confidence in that association (high, moderate, low, very low). Readers will note that this means it was possible to arrive at a conclusion of being highly confident in an inconclusive result, which holds meaning for establishing research priorities but less so for clinical practice.

Most SRs did not attempt to stratify the prognostic ability of a variable by outcome. This is understandable considering that there is little to no consensus on the most appropriate outcome to measure in prognostic research on neck pain [20]. Further, Walton and colleagues [6] attempted to evaluate the magnitude of prognostic effect between symptom-related outcomes and disability-related outcomes using meta-analysis, and showed that the magnitude of the effect was similar in almost all cases, with older age being the only notable exception. However, two SRs did present their summarized results stratified by type of outcome [5, 16]. In most cases the magnitude of association was consistent across outcomes, but where it differed, the magnitude entered into the database was the best representation of the overall reported magnitude. For example, if a predictor showed a strong association with one outcome and a limited association with another, the strength of the association for that predictor overall was described in the database as moderate. This happened in only 7 of the 239 different summary statements extracted, which are denoted in the supplementary tables.

RESULTS

Study Selection

Fig. (1) presents the results of the literature search and screening process. After applying inclusion criteria, 16 SRs were retained. During the data extraction process, we determined that 1 SR that was described as being systematic did not in fact meet our criteria for a systematic SR on prognosis as described in the Methods [21]. Another [22] did not focus specifically on prognosis in neck pain to an extent that it provided any relevant information for this OvR. These two SRs were therefore delegated to supporting evidence only. In order to avoid giving double credit to a single predictor, the updated meta-analysis of Walton and colleagues [24] and the original 2009 paper [6] were considered the same review for the purposes of this OvR. This left 13 SRs that were retained for full data extraction, 1 of which was focused solely on the course of neck pain [23], while the other 12 provided information on variables that may affect that course.

Quality Appraisal

The results of the AMSTAR appraisal process for each included review are presented in Table 1. One SR was considered high methodologic quality (low risk of bias) [6, 9, 21], 11 were considered medium quality (moderate risk of bias) [3-8, 10, 11, 17, 23, 25] and one was considered low quality (high risk of bias) [16, 26]. The majority of SRs dealt specifically with prognosis following whiplash and its

associated disorders. Other conditions were described as non-specific neck pain [11, 17], neck pain and associated disorders [25], or work-related neck pain [10] that included separate results for a sample of military personnel post cervical disc surgery. In no case were the exact same strategies for searching the literature, appraising, extracting or synthesizing the data employed, leading to several findings that were discordant between SRs even when the same or similar primary sources were included.

Prognostic Factors

A total of 133 different prognostic factors were extracted from the 12 SRs. Where multivariate analyses were used in the primary sources, most SRs used the predictors retained in the final models for establishing their levels of evidence. Otherwise, the effects of predictors were drawn from simple bivariate analyses. Table 2 presents those factors for which we have high or moderate confidence in their direction of association with outcome, either as a risk factor for poor outcome or as having no effect. These factors are also listed in the supplementary tables along with the remaining extracted factors. Brief descriptions of the results in each of the conceptual categories are described below. Readers are encouraged to consult the supplementary Tables for more detailed results (Table s1: Whiplash prognosis, Table s2: Other neck pain prognosis).

OVERALL PROGNOSIS

The SRs that synthesized the natural or clinical course of symptoms or disability in people with neck pain generally agreed that prognosis for neck pain was poor overall. Focusing specifically on whiplash, Kamper and colleagues [9] used a statistical pooling procedure to calculate a weighted mean pain intensity score of 25.3 points out of 100 and weighted mean disability score of 19 out of 100, 12 months following the initiating accident. This group also found that the majority of improvement in pain and disability occurs within the first 3 months following the accident, and plateaus considerably from that point forth. Carroll and colleagues [3] and Walton and colleagues [20] employed a more qualitative approach to synthesizing the literature. Both found a broad range of recovery rates following whiplash across primary sources. Walton [20] identified recovery rates that ranged from 16% to 99% amongst the primary studies, possibly explained by differences in operational definitions of recovery. Carroll [3] provided an overall estimate of roughly 50% of people continuing to experience some degree of neck pain 6 to 12 months following the accident. The results from the general population also provided evidence for high rates of long-term problems. Carroll and colleagues [17] reported that the balance of evidence suggests that half to three-quarters of people with neck pain will continue to report neck pain when followed up 1 to 5 years later. Hush and colleagues [23] used a statistical pooling approach to determine that the course of idiopathic neck pain was worse than previously thought, with a weighted mean pain intensity of 42 points (out of 100) when measured 12 months following onset. Disability improved at a similar rate, remaining moderate (weighted mean of 17 out of 100) at 12 months. With specific focus on work-related neck pain, Carroll and colleagues reported consistent evidence that approximately 60% of workers with neck pain

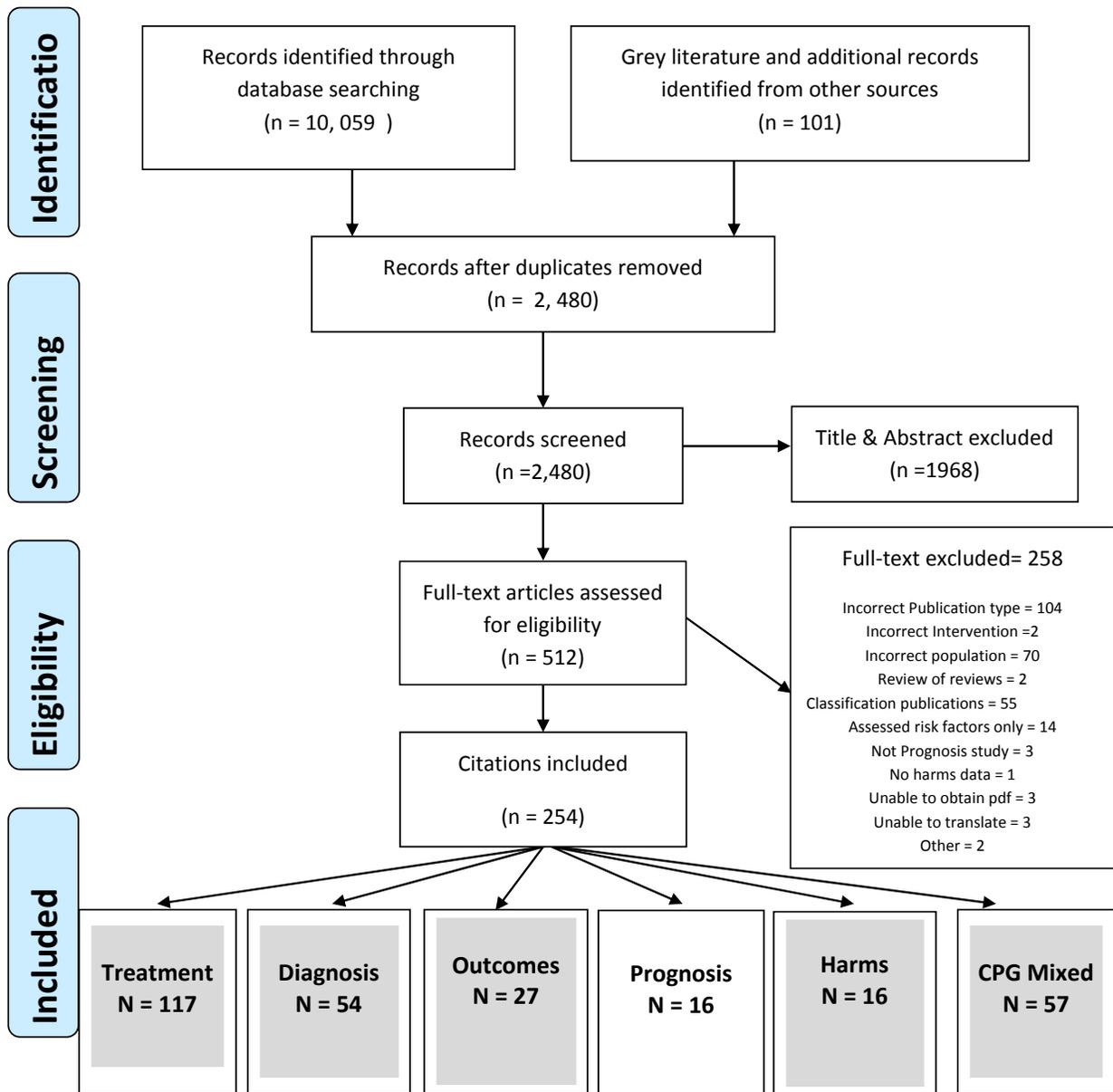


Fig. (1). The screening process for selecting the SRs to be included in the prognosis overview. Of the 16 that were considered eligible for data extraction, 1 (Tsakitidis *et al.* 2009) was deemed to be not a systematic review according to our criteria, 1 (Linton 2000) provided no useful data for the overview, and 1 (Walton *et al.* 2013) was an update of a previous meta-analysis (Walton *et al.* 2009), the two of which were considered as a single data source for this overview. This left 13 independent SRs that contributed to this overview, one of which focused solely on the course of neck pain, while 12 provided data on prognostic factors (adapted from Santaguida *et al.*, this issue).

continued to report neck pain at follow-up [10]. All SRs highlighted the challenge in synthesizing these data given the notable heterogeneity in outcomes measured across studies.

FACTORS THAT PREDICT OUTCOME IN WHIPLASH

Parameters of the Accident

Five different SRs were retrieved that reported summarized findings for the prognostic ability of accident parameters on outcomes following whiplash injury [3-6, 9, 16]. The majority of findings were in the same direction, and

suggested that, of the accident parameters evaluated in the included SRs, none had an association with outcome. Owing to the consistency of summarized findings, the existing evidence provides high confidence that the direction of the impact (rear), seating position in the vehicle (driver) and awareness of the impending collision have no effect on the outcome. We are moderately confident that whether the vehicle was moving or stationary when hit and whether the vehicles were moving at high speed also have no effect on outcome. The only exception here was change in velocity at the point of impact as measured by a crash recorder, with one SR [3] finding low evidence of a significant positive

Table 2. Predictors with Moderate or High Confidence in the Direction of their Effect on Outcome as a Result of the Triangulation Algorithm

Predictor	Condition	Primary Author (Year)	Quality of Review	Summary of Findings (From Review)	Confidence in Conclusions	Risk/ No Effect
High or Moderate Confidence as Risk Factors						
High pain intensity	Whiplash	Walton (2009) Kamper (2008) Carroll (2008) Williams (2007) Scholten-Peeters (2003)	Medium High Medium Medium Medium	<i>Strong evidence of sig. association</i> <i>Strong evidence of sig. association</i> <i>Consistent evidence of sig. association</i> <i>Moderate evidence of sig. association</i> <i>Strong evidence of sig. association</i> ¹	⊗⊗⊗⊗ High	Risk
High neck-related disability	Whiplash	Walton (2012) Kamper (2008) Carroll (2008) Williams (2007)	Medium High Medium Medium	<i>Strong evidence of sig. association</i> <i>Strong evidence of sig. association</i> <i>Strong evidence of sig. association</i> <i>Moderate evidence of sig. Association</i>	⊗⊗⊗⊗ High	Risk
Older age	Non-specific neck pain	Carroll (2008) McLean (2007)	Medium Medium	<i>Strong evidence of sig. association</i> <i>Moderate evidence of sig. association</i> ⁴	⊗⊗⊗⊗ High	Risk
Post-traumatic stress symptoms at inception	Whiplash	Kamper (2008) Williamson (2008)	High Medium	<i>Strong evidence of sig. association</i> <i>Limited evidence of sig. association</i>	⊗⊗⊗ Moderate	Risk
Catastrophizing	Whiplash	Walton (2009) Kamper (2008) Carroll (2008) Williamson (2008)	Medium High Medium Medium	<i>Moderate evidence of sig. association</i> <i>Strong evidence of sig. association</i> <i>Limited evidence of sig. association</i> Inconclusive	⊗⊗⊗ Moderate	Risk
Cold hypersensitivity/hyperalgesia	Whiplash	Kamper (2008) Williams (2007)	High Medium	<i>Moderate evidence of sig. association</i> <i>Moderate evidence of sig. association</i>	⊗⊗⊗ Moderate	Risk
History of other MSK disorders	Work-related neck pain Non-specific neck pain	Carroll (2008) McLean (2007)	Medium Medium	<i>Moderate evidence of sig. association</i> <i>Strong evidence of sig. association</i>	⊗⊗⊗ Moderate	Risk
High or Moderate Confidence as having No Effect on Outcome						
Angular deformity of the neck (scoliosis, flattened cervical lordosis)	Whiplash	Kamper (2008) Scholten-Peeters (2003)	High Medium	<i>Strong evidence of no association</i> <i>Strong evidence of no association</i>	⊗⊗⊗⊗ High	No effect
Impact direction: rear	Whiplash	Walton (2012) Kamper (2008) Carroll (2008) Scholten-Peeters (2003)	Medium High Medium Medium	<i>Strong evidence of no association</i> <i>Strong evidence of no association</i> <i>Strong evidence of no association</i> <i>Strong evidence of no association</i> <i>Strong evidence of no association</i>	⊗⊗⊗⊗ High	No effect
Seating position: driver	Whiplash	Walton (2009) Kamper (2008) Carroll (2008)	Medium High Medium	<i>Strong evidence of no association</i> <i>Strong evidence of no association</i> <i>Strong evidence of no association</i>	⊗⊗⊗⊗ High	No effect
Aware of impending collision	Whiplash	Walton (2009) Kamper (2008) Carroll (2008) Scholten-Peeters (2003)	Medium High Medium Medium	<i>Strong evidence of no association</i> <i>Strong evidence of no association</i> <i>Strong evidence of no association</i> <i>Inconclusive</i>	⊗⊗⊗⊗ High	No effect
Head rest in place	Whiplash	Walton (2009) Kamper (2008) Carroll (2008)	Medium High Medium	<i>Strong evidence of no association</i> <i>Strong evidence of no association</i> <i>Strong evidence of no association</i>	⊗⊗⊗⊗ High	No effect
Older age ²	Whiplash	Walton (2009) Kamper (2008) Carroll (2008) Scholten-Peeters (2003)	Medium High Medium Medium	<i>Moderate evidence of no association</i> ³ <i>Strong evidence of no association</i> <i>Inconclusive</i> <i>Strong evidence of no association</i>	⊗⊗⊗ Moderate	No effect
Vehicle stationary when hit	Whiplash	Walton (2009) Kamper (2008) Scholten-Peeters (2003)	Medium High Medium	<i>Strong evidence of no association</i> <i>Moderate evidence of no association</i> <i>Inconclusive</i>	⊗⊗⊗ Moderate	No effect
Regular physical activity	Non-specific neck pain Non-specific neck pain	Carroll (2009) McLean (2007)	Medium Medium	<i>Moderate evidence of no association</i> <i>Strong evidence of no association</i>	⊗⊗⊗ Moderate	No effect

1: Scholten-Peeters and colleagues were the only authors to separate the effects of pain intensity between the outcomes of pain (strong evidence) and disability (limited evidence). All other authors combined outcomes.

2: Walton and colleagues defined 'older' age as age greater than 50-55 years. Older age was not defined in the other reviews.

3: Walton and colleagues stratified the effect of older age, defined as age over 50, by outcome. For symptom-based outcomes, they found near-significant evidence of an association. For disability-based outcomes, they found strong evidence for no effect. The moderate evidence of no effect is the combined level considering these two outcomes.

4: McLean and colleagues synthesized the effect of older age across 3 different types of outcome: recovery (limited evidence of significant association), disability (moderate evidence of significant association) and symptoms (strong evidence of significant association). The indicator of *moderate* in the table is the best indicator of the overall association with the 3 types of outcome.

association (higher velocity change, greater risk of poor outcome). When velocity change was reported by the patient instead, the results of two SRs provided low confidence of no association with outcome [3, 5]. Overall, this category provided the greatest confidence in the direction of associations, but it should be noted that in the majority of primary studies included, accident parameters were self-reported rather than objectively recorded.

Psychological and Behavioural Factors

Five SRs reported summarized findings for the effect of psychological or behavioral factors on outcomes after whiplash [3, 5, 6, 8, 9]. The balance of results provided moderate confidence that early elevated post-traumatic stress symptoms at inception and highly catastrophic beliefs about pain are significant risk factors for poor outcome. A moderate pooled effect size (OR 3.77, 95%CI 1.33-10.74) was reported for high catastrophizing [24]. All other factors, including anxiety, depression, personality traits or coping behaviours provided inconclusive, low or very low confidence in their association with outcome. It is possible that a time factor may be affecting results, notably in the case of coping strategies for which Carroll and colleagues [3] found limited evidence of no association when strategies were measured within a few days of the accident, but limited evidence of a significant risk from passive coping strategies when captured in the subacute stage. Conversely, Kamper and colleagues [9] found strong evidence (3 of 4 primary studies) of a significant risk from passive coping strategies but did not describe results in terms of time from injury. Only one primary source was consistent between the two SRs [27] which may explain the discrepancy.

Self-Reported Symptoms or Interference at Inception

Five SRs reported summarized findings for the effect of early reports of pain, symptoms or disability on outcomes after whiplash [3-7, 9]. Owing to the consistency of findings, the balance of evidence provided high confidence that higher pain intensity and self-reported disability at inception were predictors of poorer outcome. Pain intensity in particular is consistently reported as a strong predictor, with one pooled effect size reported for pain intensity of 5.5/10 (55/100) or greater (OR 5.61, 95%CI 3.74-8.43) [24]. An NDI score of greater than 15/50 points at baseline provided a large pooled effect when disability was the predicted outcome (OR 42.18) but with very broad confidence limits that limit confidence in the point estimate (95%CI 7.37 to 241.3) [24]. Beyond those two indicators however, the remaining factors provide inconclusive, low or very low confidence in their association with outcome. Perhaps more than the other categories, this one highlighted the differences between SRs in terms of strategies and the subsequent conclusions drawn. As an illustrative example, 4 SRs evaluated the effect of number of different symptoms or areas of the body in pain as a prognostic factor, two of which were published in the same year. Kamper and colleagues [9] found inconclusive evidence of an association with outcome, while Carroll and colleagues [3] reported strong evidence of a significant association. Deeper exploration of the results and supplemental tables of these two SRs revealed that the two primary papers that informed the results of the Kamper [9] SR were excluded from the Carroll [3] SR, while all 4

primary sources that informed the Carroll [3] SR were excluded from the Kamper [9] SR.

This category also provided evidence of a possible effect from time-to-follow-up on the prognostic value of some factors. Kamper and colleagues [9] found moderate evidence of a significant association between patient-reported radicular symptoms at inception and risk of poor outcome at follow-up. However, when stratified by time-to-follow-up, their primary sources found no evidence of an association with outcome when captured less than 6 months following the accident, but 2 of 3 primary sources [28, 29] found a significant association when outcome was captured 6 months or longer following the accident. The other SRs did not make this time-related distinction which may account for the inconsistent findings.

Biological and/or Clinical Assessment

Five SRs presented summarized findings for the association between observational clinical or diagnostic findings at inception and outcome [3, 5-7, 9]. The balance of evidence provides moderate confidence that cold hypersensitivity/hyperalgesia is a risk factor for poor outcome, with a low to moderate association but no pooled effect size reported. The synthesis also provided high confidence that angular deformities of the neck (e.g. scoliosis, flattened cervical lordosis) have no effect on outcome. Consistently inconclusive findings were reported for each of the effect of reduced cervical range of motion, morphological changes on diagnostic imaging, and body mass index, suggesting the need for greater standardization of these variables.

Medicolegal Context

Three SRs presented summarized findings specific to the medicolegal context within which the injury occurred [3-5, 16]. The balance of the findings did not provide high or moderate confidence for any of the associations between the 3 medicolegal factors (compensation system, receiving compensation, lawyer involvement) and outcome. While there were no inconclusive findings in this category, the strength of the evidence at the time the SRs were conducted did not allow for strong conclusions to be drawn.

Other Social Influences

Three SRs presented summarized findings for the effect of other social influences (outside of the medicolegal context) on outcome after whiplash [4, 5, 8, 16]. These included the type of work, 'psychosocial' work factors and social support. The strength of the evidence included in each of the 3 SRs prevented the drawing of conclusions with anything greater than very low confidence.

Demographics

Four SRs presented summarized results for the effects of demographic variables (sex, age, education) on outcomes after whiplash [3-6, 9]. The balance of evidence as reported in the SRs provided moderate confidence that age had no effect on outcome, but this finding was not universal. Only Walton and colleagues [6] attempted to define 'older' age as age greater than 50 years in their meta-analysis. They found that the effect of age on outcome may vary by type of outcome, reporting strong evidence of no effect when

disability was the outcome, while finding a near-significant positive effect when symptoms (pain) was the outcome.

The effect of sex on outcome was inconclusive, with one high-quality SR [9] finding strong evidence of no effect, one medium quality SR [3] finding inconclusive results, and one meta-analysis [6] providing moderate evidence of a significant risk for females compared to males. Deeper exploration of each SR revealed that, of the primary sources reviewed by Kamper and colleagues [9], only 2 of 17 suggested a significant risk for females. In contrast, 7 of the 11 primary sources reviewed by Carroll and colleagues [3] found significance, with the other 4 finding significant bivariate associations only when sex was evaluated in isolation but not as part of a multivariate model. The meta-analytic approach of Walton and colleagues [6] found a small but significant effect only when findings from 11 primary sources were statistically pooled (OR 1.64, 95%CI 1.27 to 2.12), but only 5 of 11 found female gender to be a significant risk factor when analyzed in isolation from other variables.

The effect of education was similarly inconsistent: Kamper and colleagues [9] reported that 2 of 4 primary sources suggested a significant risk of poor outcome amongst those with lower education, Carroll and colleagues [3] found evidence of a significant risk in 2 of 3 primary sources while the third suggested a protective effect of lower education. As was the case with female sex, Walton and colleagues [6] found lower education, defined as less than post-secondary, was a significant risk factor only after the statistical pooling procedure (OR 2.00, 95%CI 1.60 to 2.51). Again, clear differences in the strategies employed to search, appraise, extract and synthesize the literature led to different findings across SRs in the same area.

Treatment-Related Factors

Four SRs evaluated the effect that treatment-related factors (type, frequency or duration of treatment) had on outcome after whiplash [3-5, 16]. Interestingly, none of the variables in this category had been summarized in more than one SR and only 1 of the 3 SRs was published within the past 5 years. Therefore, the synthesis framework provided only very low confidence in the effect of the treatment-related factors, preventing any firm conclusions from being drawn regarding the association between treatment and outcome. Where significant associations did exist, the results generally suggested greater use of medical or rehabilitation services early was associated with poorer long-term outcomes. However, multivariate models were rare, rendering any discussion of causal mechanisms between treatment and outcome inappropriate.

Pre-Injury History

Seven SRs presented at least one summarized finding for the effect of pre-injury history on outcomes after whiplash [3-9, 16]. The majority of findings in this category were inconclusive. As was the case for the parameters of the accident, these variables were almost universally collected by self-report in the primary sources, presenting a strong possibility of biased estimates.

FACTORS THAT PREDICT OUTCOME FROM OTHER CAUSES OF NECK PAIN

Five SRs presented prognostic data specific to neck pain conditions other than whiplash [10, 11, 17, 25, 30] (Table s2). These conditions included neck pain in workers, 'neck pain and associated disorders', non-specific neck pain, or post-surgical neck pain in military personnel. From all 5 SRs, 37 independent predictors could be extracted. In the majority of cases, each predictor was evaluated in only 1 SR and was found to have limited evidence of risk for poor outcome, providing low or very low confidence in their direction of association. Only 2 factors provided high or moderate confidence in their ability to predict a poor outcome. The first was a history of other musculoskeletal disorders (other than neck pain) prior to the current episode of neck pain, for which two SRs found moderate [10] or strong evidence [11] that it was a risk factor for a poor outcome in work-related or non-specific neck pain, respectively. Two SRs found strong [17] or moderate [11] evidence for older age (not defined) as a predictor of poor outcome in non-specific neck pain, while moderate evidence for no effect was found for work-related neck pain [10]. In no case was an effect size reported. While confidence in the association is low, engaging in physical exercise as a lifestyle habit prior to onset of neck pain may have a protective effect against long-term problems, the only predictor to be reported as such. As was the case for whiplash-related neck pain, these questions are generally captured through self-report, and despite some consistency in the evidence, these methods are prone to recall or social desirability bias.

DISCUSSION

The results of an 'overview' of reviews (systematic review of systematic reviews) suggested that the prognosis of neck pain of various causes is generally poor and there are relatively few factors that allow high or moderate confidence in their use as predictors of outcome. We used a decision algorithm that favoured recent, high- to medium-quality SRs to determine the association (risk, no effect, inconclusive) between 133 predictors and a broad operationalization of 'outcome' that included one or more type of measure to assess pain, disability, work status, time-to-claim-closure, or some combination of these. The algorithm also provided an indication of our confidence in the direction and strength of these associations, using categories adapted from the GRADE working group (high, medium, low, very low). A notable outcome of this exercise was the frequency with which two or more different authorship groups reached quite different conclusions regarding the same predictor, highlighting the impact of different strategies for literature search, appraisal, extraction and synthesis.

High baseline pain intensity and to a slightly lesser extent high baseline self-reported neck disability are new universal predictors of prolonged recovery. Angular deformities of the neck, along with several parameters of the accident itself (direction of impact, seating position, awareness of collision, use of a head rest) are consistently found to have no effect on recovery especially when self-reported by the patient. In

terms of non-whiplash-related neck pain, we are moderately confident that a past history of 'other' musculoskeletal disorders (other than neck, shoulder, headache or low back pain) is a risk factor for prolonged recovery; that older age may prolong recovery from non-specific neck pain; and that regular physical activity has no clear effect on outcome. Catastrophizing, cold hyperalgesia and acute post-traumatic stress response round out the remaining risk factors for which the evidence provides moderate confidence in their prognostic ability, but more research with consistent predictors, duration of follow-up and outcomes is required for firm conclusions.

Many factors have been evaluated only once or have conflicting results, and hence provide low or very low confidence in their association with outcome. There are also several inconclusive findings reported in the supplementary tables, suggesting that these factors may or may not be predictive but require further study. These findings support the need for further large cohort studies that assess prognostic factors in an accurate and comprehensive manner to provide definitive estimates of the effects of these potentially useful predictors. With specific focus on the conduct of systematic reviews, the number of conflicting or inconclusive results also suggests inconsistent methodologies that provide clinicians or policy-makers with very different information depending on the SR chosen. For example, the evidence for the predictor '*General psychological distress at baseline*' has been synthesized by three independent groups. One high-quality SR [9] found strong evidence for a significant association between the magnitude of acute general psychological distress, broadly defined, and follow-up outcome, also broadly defined. Conversely, two medium quality SRs [5, 8] found moderate or strong evidence of no association with outcome, despite two of the conflicting SRs having been published in the same year. Of interest here is that the approach to data synthesis employed by Kamper and colleagues [9] led to findings of a significant association between psychological distress and outcome being drawn from the primary sources of Hendriks and colleagues [28] and Olsson and colleagues [31]. The same two primary sources were evaluated by Williamson and colleagues [8], whose data synthesis algorithm suggested that these same two sources provided no evidence of a significant association. It is not our intention to comment on the validity of either data synthesis approach, rather disparate findings such as these warrant caution in interpreting even systematic reviews, and highlight the value of periodic overviews such as the one presented herein. This factor is also just one example of the impact that differences in operationalization of either the predictor (general psychological distress) or the outcomes between authorship groups can have on reported findings. The decision algorithm as used in our overview gives heavier weighting to the results of recent, higher quality SRs. As a result our triangulation exercise resulted in very low confidence of a significant risk, but the results from each of the individual SRs provide a very different picture. Given the frequency with which policy makers rely on systematic reviews for establishing the state of evidence in an area, findings such as these demand caution and provide rationale for considering more than one source when policy decisions are to be made.

While conflicting results may be difficult to fully explain, consistent results for predictors across SRs, despite different methodologies, provide greater confidence in their association. High pain intensity has consistently shown a strong association with poor outcomes after WAD, but not so in non-specific neck pain, for which a single medium-quality SR was included that provided inconclusive results [11]. Even in patients with WAD, a finding of high intensity neck pain does little to explain the mechanism. It has long been recognized that the experience of pain is a multifactorial phenomenon, influenced by sensory, evaluative and affective domains [32]. Since pain intensity is most commonly captured through a 0-10 or 0-100 rating scale, it is impossible to determine which, if any, one of the domains of pain experience should be the target of intervention. Further, recent models have encouraged mechanism-based assessment of pain [33] which appears to be especially relevant in WAD, for which clear tissue-based pathology can rarely explain the magnitude of symptoms [34]. Evidence exists to support an understanding of some manifestations of WAD as a neuropathic pain condition [35], or as a consequence of some neuroplastic change at the level of central nociceptive processing [36]. Evidence also exists, and continues to build, for the role of acute post-traumatic stress reactions as a predictor of poor outcome, and the relationship of such reactions with objective signs of nociceptive sensory dysfunction [37]. The results of the current overview would suggest that simple assessment of pain intensity is a valuable tool in establishing a prognosis following acute WAD, and also suggests that researchers need not dedicate further resources to establishing this relationship. Rather, resources should be dedicated toward evaluating the influences on, or mechanisms of, the acute pain experience to provide clinicians with greater guidance in clinical decision making about how to deal with the subset of patients who have an adverse prognosis as indicated by high baseline pain. The same argument can be made for neck-related disability, which is often reported as a composite score across several symptom and function domains (usually including pain intensity). As with pain intensity, an aggregate score on a disability scale may provide value from a prognostic standpoint, but does little to guide clinical decisions. Individual items on a multidimensional disability scale may be clinically useful for guiding treatment decisions, but rarely are the responses to individual items reported in the literature and their unique prognostic ability is largely unknown.

The balance of evidence as included in the SRs would suggest that self-reported constructs, such as pain intensity, disability, psychological distress or coping strategies are stronger predictors of outcome than are the more observational signs such as structural pathology on diagnostic imaging, cervical range of motion or angular deformities. It is tempting to assume then, that the cognitive aspects of the experience of neck pain are more important to its experience and subsequent recovery than are the physical aspects. However, we urge caution in this interpretation. Readers should recognize that the majority of operational definitions for recovery are also self-reported, most commonly being heavily weighted towards pain or disability [20]. Statistically, when attempting to identify a set of predictors that are able to explain the greatest variance in the

outcome, as is the case in the often employed linear regression approach, it should come as no surprise that self-reported predictors are better able to explain that variance than are biological indicators. This is especially true when those predictors are captured on the same scale as the outcome, while clinical or biological signs are captured on very different scales, and often with limited statistical distribution. Even many quantitative sensory tests are best viewed as self-report measures as they rely on cognitive processes to determine what the patient considers to be painful. A further consideration is that predictors are more likely to demonstrate value when measured using tools with sound psychometric or clinometric properties (reliable and valid). On balance, the literature provides greater evidence of sound properties for psychological or screening questionnaires than for clinical tests. As an illustrative example, two primary studies in whiplash have shown that a rigorously-developed and validated protocol for measuring cervical range of motion was the strongest predictor of outcome, even when evaluated in the same multivariate model as self-report measures [38,39]. It is rare that clinical tests can claim such strong measurement properties. Additional rigorously-developed objective, observational clinical or biological tests might provide different insights into the risk and mechanisms of transition from acute to chronic neck pain. The development of sound tools is a reasonable direction for further research. Since there are many potential structures and processes that are affected in WAD, the absence of comprehensive structural and physiologic diagnostic regimens may mean that the sequelae of undocumented impairments on these domains are manifested through higher pain and self-report. Without such diagnostic tools, the physiologic and psychological components of neck disorders can be difficult to disentangle.

The effect of the medicolegal environment on outcome continues to be debated in the scientific and lay literature [40], and our overview provides little clarity. Very few studies have addressed this issue specifically, likely due to the challenges of doing so in a scientifically sound manner. Our results provide low confidence that a no-fault insurance system may provide some protective effect compared to a tort system. Even the effect on outcome of retaining a lawyer shortly after the accident is unclear, with one medium-quality SR providing moderate or limited evidence of an association with outcome [3, 4], and one low quality SR providing consistent evidence [16]. Intuitively and anecdotally, being involved in litigation shortly after an accident or injury would be associated with prolonged recovery. As is the case with pain intensity however, the mechanisms are unclear. Whether it is being involved in the largely adversarial medicolegal environment that prolongs outcome, or that people who perceive themselves as less likely to recover are more likely to seek litigation or compensation, is unclear. This issue of causation and reverse-causation is beyond the scope of this paper, but readers are directed to Spearing and Connelly [40] and Carroll and colleagues [41] for a deeper discussion.

A key finding that has come to light as a result of this OvR is the conclusion that finding evidence to support simple bivariate associations between a single predictor and an outcome is not difficult when the study is of adequate methodological rigour. However, clinicians and researchers

will recognize that it is highly uncommon to find a single risk factor in an individual patient, rather multiple risk factors are present that likely interact with each other in complex ways to affect the course of recovery. The body of knowledge in the field is starkly insufficient for explaining how the effects of 2, 3, or more, risk factors interact to influence outcome. Exploratory approaches to identifying risk factors, such as multiple linear regression, may be useful in the early stages of research. However, we suggest that more complex confirmatory testing of *a priori* established theoretical multivariate models is required in order to fit knowledge from this field into the complexities of clinical practice and true human interaction. Suggestions here are to move from simple correlational analyses to structural equation modeling or a subtype (e.g. latent class growth curve analysis) to identify different trajectories from acute injury and subsequent algorithms for predicting those trajectories. Integrative prognostic models have been proposed previously that could become testable with some additional work, or have already been tested with promising early findings [42-45]. Researchers are encouraged to consider more advanced modeling techniques in the conduct of future research in this area.

There are two key limitations that must be considered when evaluating the clinical usefulness of our results. The first is that, by the nature of an overview, we limited our data extraction to what was presented in previous SRs. While the exclusion of narrative reviews or commentaries/editorials limits the risk of bias somewhat, readers should remember that the information we extracted had already been filtered once by a previous group of reviewers. As such, our data would have been subject to the same biases or methodologic weaknesses as were present in the included primary SRs. In an effort to mitigate this concern, we used a well-established instrument, the AMSTAR methodological quality checklist [12], then constructed a conservative algorithm that gave more weighting to more recent and higher-quality SRs, and finally, we assembled an authorship group that had representation from many of the included primary SRs. While these steps may safeguard against bias to a degree, the nature of an overview is that the data extracted are only as good as what was presented in the primary SRs, and those data are only as good as what was drawn from the primary studies. This should be considered when interpreting our results.

We feel this systematic analysis was valuable insofar as many of the SRs included here continue to be accessed by clinicians and policy-makers when it comes to issues of establishing prognosis in acute WAD or other neck pain. The decision algorithm, establishing the level of evidence across SRs, is a novel addition to the existing pool of literature, and should serve as a solid start point for new research in the area. More recently, evidence has largely continued to support the value of acute post-traumatic stress reactions [46], catastrophizing or pain-related beliefs [47], quantitative sensory testing [48] and expectation [49] as predictors of outcome. Efforts have been made to construct standardized risk screening tools [50] or to identify biological correlates of neck pain [51], each of which are in their infancy but have shown early promise in furthering this field. The prognostic value of thermal [36] or mechanical [48, 52] pain threshold

continues to be evaluated, and may become valuable proxies for disordered nociceptive processing.

In summary, we have conducted systematic overview of systematic SRs to identify consistent findings and establish the level of confidence in the field of prognosis after neck pain, and to organize the current body of evidence upon which future systematic reviews can build. The majority of this work has been conducted in whiplash-associated disorder, possibly owing to the ease with which time from injury to inception can be established and the potential to quantify the magnitude of the event. Self-reported constructs, especially high pain intensity and neck-related disability, have well-established evidence for their value as predictors of poor outcome. Efforts should now be directed towards deeper exploration of the pain or disability experience, including the biopsychosocial domains of pain and disability and mechanisms behind their genesis. Consensus on important outcomes, establishment of valid, reliable and useful clinical and biological markers of dysfunction, and identification of the most parsimonious set of variables through advanced multivariate modeling techniques are all ripe fields for future study. The large number of inconclusive or low-to-very-low confidence findings suggests there is still a considerable amount of work to be done in the field of prognosis after neck pain.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

ICON is a multi-disciplinary collaborative group that includes scientist-authors (listed below) and support staff (Margaret Lomotan) that conduct knowledge synthesis and translation aimed at reducing the burden of neck pain.

The ICON authors that provided direction of the project and reviewed the findings/manuscript include (in alphabetical order): Gert Bronfort, Norm Buckley, Lisa Carlesso, Linda Carroll, Pierre Côté, Jeanette Ezzo, Paulo Ferreira, Tim Flynn, Charlie Goldsmith, Anita Gross, Ted Haines, Jan Hartvigsen, Wayne Hing, Gwendolyn Jull, Faith Kaplan, Ron Kaplan, Helge Kasch, Justin Kenardy, Per Kjær, Janet Lowcock, Joy MacDermid, Jordan Miller, Margareta Nordin, Paul Peloso, Jan Pool, Duncan Reid, Sidney Rubinstein, Lina Santaguida, Anne Söderlund, Natalie Spearing, Michele Sterling, Grace Szeto, Robert Teasell, Arianne Verhagen, David M. Walton, Marc White.

This work was funded through a Knowledge Translation grant (FRN: KRS-102084) from the Canadian Institutes of Health Research. None of the authors received any direct compensation for their work on this project.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's web site along with the published article.

REFERENCES

[1] Croft PR, Lewis M, Papageorgiou AC, *et al.* Risk factors for neck pain: a longitudinal study in the general population. *Pain* 2001; 93(3): 317-25.
 [2] Hogg-Johnson S, van der Velde G, Carroll LJ, *et al.* The burden and determinants of neck pain in the general population: results of

the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *J Manipulative Physiol Ther* 2009; 32(Suppl 2): S46-60.
 [3] Carroll LJ, Holm LW, Hogg-Johnson S, *et al.* Course and prognostic factors for neck pain in whiplash-associated disorders (WAD): results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine* 2008; 33(Suppl 4): S83-92.
 [4] Cote P, Cassidy JD, Carroll L, Frank JW, Bombardier C. A systematic review of the prognosis of acute whiplash and a new conceptual framework to synthesize the literature. *Spine* 2001; 26(19): E445-58.
 [5] Scholten-Peeters GG, Verhagen AP, Bekkering GE, *et al.* Prognostic factors of whiplash-associated disorders: a systematic review of prospective cohort studies. *Pain* 2003; 104(1-2): 303-22.
 [6] Walton DM, Pretty J, MacDermid JC, Teasell RW. Risk factors for persistent problems following whiplash injury: results of a systematic review and meta-analysis. *J Orthop Sports Phys Ther* 2009; 39(5): 334-50.
 [7] Williams M, Williamson E, Gates S, Lamb S, Cooke M. A systematic literature review of physical prognostic factors for the development of Late Whiplash Syndrome. *Spine* 2007; 32(25): E764-80.
 [8] Williamson E, Williams M, Gates S, Lamb SE. A systematic literature review of psychological factors and the development of late whiplash syndrome. *Pain* 2008 Mar; 135(1-2): 20-30.
 [9] Kamper SJ, Rebbeck TJ, Maher CG, McAuley JH, Sterling M. Course and prognostic factors of whiplash: A systematic review and meta-analysis. *Pain* 2008 ; 138(3): 617-29.
 [10] Carroll LJ, Hogg-Johnson S, Cote P, *et al.* Course and prognostic factors for neck pain in workers: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine* 2008; 33(Suppl 4): S93-100.
 [11] McLean SM, May S, Moffett J, Sharp DM, Gardiner E. Prognostic factors for progressive non-specific neck pain: a systematic review. *Phys Ther Rev* 2007; 12(3): 220.
 [12] Shea BJ, Grimshaw JM, Wells GA, *et al.* Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007; 7: 10.
 [13] Shea BJ, Hamel C, Wells GA, *et al.* AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol* 2009; 62(10): 1013-20.
 [14] Popovich I, Windsor B, Jordan V, Showell M, Shea B, Farquhar CM. Methodological quality of systematic reviews in subfertility: a comparison of two different approaches. *PLoS ONE* 2012; 7(12): e50403.
 [15] Whitlock EP, Lin JS, Chou R, Shekelle P, Robinson KA. Using existing systematic reviews in complex systematic reviews. *Ann Intern Med* 2008; 148(10): 776-82.
 [16] McClune T, Burton AK, Waddell G. Whiplash associated disorders: a review of the literature to guide patient information and advice. *Emerg Med J* 2002; 19(6): 499-506.
 [17] Carroll LJ, Hogg-Johnson S, van der Velde G, *et al.* Course and prognostic factors for neck pain in the general population: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *J Manipulative Physiol Ther* 2009; 32(Suppl 2): S87-96.
 [18] Flick U. Triangulation in qualitative research. In: Flick U, von Kardorff E, Steinke I, Eds. *A companion to qualitative research*. London: Sage 2004; pp. 178-83.
 [19] Guyatt GH, Oxman AD, Vist GE, *et al.* GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *Br Med J* 2008; 336(7650): 924-6.
 [20] Walton D. A review of the definitions of recovery used in prognostic studies on whiplash using an ICF framework. *Disabil Rehabil* 2009; 31(12): 943-57.
 [21] Tsakitzidis G, Remmen R, Peremans L, Van Royen P. Non-specific neck pain: diagnosis and treatment. *KCE reports* 119c, 2009.
 [22] Linton SJ. A review of psychological risk factors in back and neck pain. *Spine* 2000; 25(9): 1148-56.
 [23] Hush JM, Lin CC, Michaleff ZA, Verhagen A, Refshauge KM. Prognosis of acute idiopathic neck pain is poor: a systematic review and meta-analysis. *Arch Phys Med Rehabil* 2011; 92(5): 824-9.
 [24] Walton DM, Macdermid JC, Giorgianni AA, Mascarenhas JC, West SC, Zammit CA. Risk Factors for Persistent Problems

- Following Acute Whiplash Injury: Update of a Systematic Review and Meta-Analysis. *J Orthop Sports Phys Ther* 2013; 43(2): 31-43.
- [25] Cassidy JD, Cote P. Is it time for a population health approach to neck pain? *J Manipulative Physiol Ther* 2008; 31(6): 442-6.
- [26] Wainner RS, Gill H. Diagnosis and nonoperative management of cervical radiculopathy. *J Orthop Sports Phys Ther* 2000; 30(12): 728-44.
- [27] Carroll LJ, Cassidy JD, Cote P. The role of pain coping strategies in prognosis after whiplash injury: passive coping predicts slowed recovery. *Pain* 2006; 124(1-2): 18-26.
- [28] Hendriks EJ, Scholten-Peeters GG, van der Windt DA, Neeleman-van der Steen CW, Oostendorp RA, Verhagen AP. Prognostic factors for poor recovery in acute whiplash patients. *Pain* 2005; 114(3): 408-16.
- [29] Radanov BP, Sturzenegger M, De Stefano G, Schnidrig A. Relationship between early somatic, radiological, cognitive and psychosocial findings and outcome during a one-year follow-up in 117 patients suffering from common whiplash. *Br J Rheumatol* 1994; 33(5): 442-8.
- [30] Campbell P, Wynne-Jones G, Dunn KM. The influence of informal social support on risk and prognosis in spinal pain: a systematic review. *Eur J Pain* 2011; 15(5): 444.e1-444.14.
- [31] Olsson I, Bunketorp O, Carlsson SG, Styf J. Prediction of outcome in whiplash-associated disorders using West Haven-Yale Multidimensional Pain Inventory. *Clin J Pain* 2002; 18(4): 238-44.
- [32] Melzack R. From the gate to the neuromatrix. *Pain* 1999; Suppl 6: S121-6.
- [33] Woolf CJ, Bennett GJ, Doherty M, *et al.* Towards a mechanism-based classification of pain? *Pain* 1998; 77(3): 227-9.
- [34] Nordin M, Carragee EJ, Hogg-Johnson S, *et al.* Assessment of neck pain and its associated disorders: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine* 2008; 33(Suppl 4): S101-22.
- [35] Sterling M, Pedler A. A neuropathic pain component is common in acute whiplash and associated with a more complex clinical presentation. *Man Ther* 2009; 14(2): 173-9.
- [36] Sterling M. Differential development of sensory hypersensitivity and a measure of spinal cord hyperexcitability following whiplash injury. *Pain* 2010; 150(3): 501-6.
- [37] Sterling M, Kenardy J. The relationship between sensory and sympathetic nervous system changes and posttraumatic stress reaction following whiplash injury—a prospective study. *J Psychosom Res* 2006; 60(4): 387-93.
- [38] Kasch H, Bach FW, Jensen TS. Handicap after acute whiplash injury: a 1-year prospective study of risk factors. *Neurology* 2001; 56(12): 1637-43.
- [39] Kasch H, Qerama E, Kongsted A, Bendix T, Jensen TS, Bach FW. Clinical assessment of prognostic factors for long-term pain and handicap after whiplash injury: a 1-year prospective study. *Eur J Neurol* 2008; 15(11): 1222-30.
- [40] Spearing NM, Connelly LB. Whiplash and the compensation hypothesis. *Spine (Phila Pa 1976)* 2011; 36(Suppl 25): S303-8.
- [41] Carroll LJ, Connelly LB, Spearing NM, Cote P, Buitenhuis J, Kenardy J. Complexities in understanding the role of compensation-related factors on recovery from whiplash-associated disorders: discussion paper 2. *Spine (Phila Pa 1976)* 2011; 36(Suppl 25): S316-21.
- [42] Guzman J, Hurwitz EL, Carroll LJ, *et al.* A new conceptual model of neck pain: linking onset, course, and care: the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine* 2008; 33(Suppl 4): S14-23.
- [43] McLean SA, Clauw DJ, Abelson JL, Liberzon I. The development of persistent pain and psychological morbidity after motor vehicle collision: integrating the potential role of stress response systems into a biopsychosocial model. *Psychosom Med* 2005; 67(5): 783-90.
- [44] Sullivan MJ, Thibault P, Simmonds MJ, Milioto M, Cantin AP, Velly AM. Pain, perceived injustice and the persistence of post-traumatic stress symptoms during the course of rehabilitation for whiplash injuries. *Pain* 2009; 145(3): 325-31.
- [45] Sterling M, Hendrikz J, Kenardy J, *et al.* Assessment and validation of prognostic models for poor functional recovery 12months after whiplash injury: A multicentre inception cohort study. *Pain* 2012; 153(8): 1727-34.
- [46] Sterling M, Hendrikz J, Kenardy J. Compensation claim lodgement and health outcome developmental trajectories following whiplash injury: A prospective study. *Pain* 2010; 150(1): 22-8.
- [47] Bostick GP, Carroll LJ, Brown CA, Harley D, Gross DP. Predictive capacity of pain beliefs and catastrophizing in Whiplash Associated Disorder. *Injury* 2012; [Epub ahead of print].
- [48] Walton D, MacDermid J, Nielson W, Teasell R, Reese H, Levesque L. Pressure pain threshold testing demonstrates predictive ability in people with acute whiplash. *J Orthop Sports Phys Ther* 2011; 41(9): 658-65.
- [49] Ozegovic D, Carroll LJ, David Cassidy J. Does expecting mean achieving? The association between expecting to return to work and recovery in whiplash associated disorders: a population-based prospective cohort study. *Eur Spine J* 2009; 18(6): 893-9.
- [50] Kasch H, Qerama E, Kongsted A, Bach FW, Bendix T, Jensen TS. The risk assessment score in acute whiplash injury predicts outcome and reflects biopsychosocial factors. *Spine (Phila Pa 1976)* 2011; 36(Suppl 25): S263-7.
- [51] Elliott J, Pedler A, Kenardy J, Galloway G, Jull G, Sterling M. The temporal development of fatty infiltrates in the neck muscles following whiplash injury: an association with pain and posttraumatic stress. *PLoS ONE* 2011; 6(6): e21194.
- [52] Kasch H, Qerama E, Kongsted A, Bach FW, Bendix T, Jensen TS. Deep muscle pain, tender points and recovery in acute whiplash patients: a 1-year follow-up study. *Pain* 2008; 140(1): 65-73.

Received: December 21, 2012

Revised: January 2, 2013

Accepted: January 2, 2013

© Walton *et al.*; licensee *Bentham Open*.This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.